



Published in final edited form as:

*Nat Biotechnol.* 2013 April ; 31(4): 342–349. doi:10.1038/nbt.2519.

## Deciphering molecular circuits from genetic variation underlying transcriptional responsiveness to stimuli

Irit Gat-Viks<sup>1,2,¶</sup>, Nicolas Chevrier<sup>1,3,4</sup>, Roni Wilentzik<sup>2</sup>, Thomas Eisenhaure<sup>1,5</sup>, Raktima Raychowdhury<sup>1</sup>, Yael Steurman<sup>2</sup>, Alex Shalek<sup>6</sup>, Nir Hacohen<sup>1,7</sup>, Ido Amit<sup>1,8,¶</sup>, and Aviv Regev<sup>1,9,¶</sup>

<sup>1</sup> Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA

<sup>2</sup> Department of Cell Research and Immunology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel

<sup>3</sup> Graduate Program in Immunology, Division of Medical Sciences, Harvard Medical School, Boston, MA 02115, USA

<sup>4</sup> FAS Center for Systems Biology, Harvard University, Cambridge, MA 02138, USA

<sup>5</sup> Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Charlestown, MA 02129, USA

<sup>6</sup> Departments of Chemistry and Chemical Biology and of Physics, Harvard University, Cambridge, MA 02138, USA

<sup>7</sup> Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Charlestown, MA 02129, USA, and Department of Medicine, Harvard Medical School, Boston, MA 02115, USA

<sup>8</sup> Department of Immunology, Weizmann Institute of Science, Rehovot, 76100, Israel

<sup>9</sup> Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02140, USA

### Abstract

Individual genetic variation affects gene expression in response to stimuli, often by influencing complex molecular circuits. Here we combine genomic and intermediate-scale transcriptional profiling with computational methods to identify variants that affect the responsiveness of genes to stimuli (responsiveness QTLs; reQTLs) and to position these variants in molecular circuit

Users may view, print, copy, download and text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: [http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

¶ To whom correspondence should be addressed: iritgv@post.tau.ac.il (IGV), ido.amit@weizmann.ac.il (IA), aregev@broad.mit.edu (AR).

#### Contributions

I.G.V., I.A. and A.R. conceived and designed the study. N.C., T.E, R.R., A.S. and I.A. conducted the experiments. I.G.V. and A.R. conceived computational methods. I. G. V., R.W. and Y.S. conceived, developed and implemented the computational methods. N. H. participated in study design and interpretation. I.G.V., I.A., and A. R. wrote the manuscript with input from all the authors.

#### Competing financial interests

The authors declare no competing financial interests.

**Accession codes.** The complete microarray data sets are available from the Gene Expression Omnibus (GSE40452).

diagrams. We apply this approach to study variation in transcriptional responsiveness to pathogen components in dendritic cells from recombinant inbred mouse strains. We identify reQTLs that correlate with particular stimuli and position them in known pathways. For example, in response to a virus-like stimulus, a *trans*-acting variant acts as an activator of the antiviral response; using RNAi, we identify Rgs16 as the likely causal gene. Our approach charts an experimental and analytic path to decipher the mechanisms underlying genetic variation in circuits that control responses to stimuli.

---

## Introduction

Genetic variation can affect molecular circuits that control mRNA levels<sup>1-4</sup>. In particular, genetic variants that affect the absolute abundance of transcripts are called expression quantitative trait loci (eQTLs), whereas variants that affect changes in transcript abundance after perturbation by external stimuli are referred to as responsiveness expression quantitative trait loci (reQTLs)<sup>4</sup>. Recent studies demonstrated the utility of testing reQTLs in the context of a single stimulus (*e.g.*, ionizing radiation) in yeast and mammals<sup>1, 4-10</sup>; some reQTLs could be detected in response to stimulus, but not using baseline mRNA abundance levels.

Although most molecular circuits can respond to multiple distinct stimuli, it remains unknown how the same reQTL would behave in response to different stimuli in the same cell type. The reQTL associated with a given gene could either affect that gene's responsiveness to all the stimuli to which the gene responds (stimulus non-specific reQTLs; **Fig. 1a, left**), or may influence gene expression in response to some but not all of the stimuli to which the gene responds (stimulus-specific reQTLs, **Fig. 1a, right**). Earlier studies monitored reQTLs only during a single stimulus and hence could not distinguish these possibilities.

Determining the stimulus-specificity of a reQTL can, in principle, allow us to position it within complex circuits. For example, consider a circuit with three input signals, two target genes, and the network connecting them (**Fig. 1b,c**). Determining a *trans*-reQTL in just one stimulus (*e.g.*,  $s_2$ , **Fig. 1b**), would position it upstream of its target gene(s) (*e.g.*,  $x_1$ , **Fig. 1b**), but cannot predict on which of the multiple paths from stimulus to target it resides (*e.g.*, both pathways 2 and 3 are possible positions, **Fig. 1b**). However, if one determines a *trans*-reQTL across a sufficiently diverse set of stimuli (*e.g.*,  $s_1$ ,  $s_2$ , and  $s_3$ , **Fig. 1b**), it can be positioned on a particular pathway in the circuit (*e.g.*, pathway 2, **Fig. 1b**). A similar rationale applies to *cis*-acting variants (**Fig. 1c**). If the genetic architecture cannot be explained in the context of the known circuitry, we can use the inconsistency to refine the circuit with new putative connections.

Here, we systematically study these questions by developing a framework, involving several new computational methods, for multi-stimulus reQTL analysis. We apply this approach to study the transcriptional responsiveness of bone-marrow-derived primary dendritic cells (DCs) to stimulation with different pathogen components, across a panel of recombinant inbred (RI) BXD mice, generated by crossing the parental C57BL/6J (B6) and DBA/2J (D2)

strains<sup>11</sup>. We find that stimulation-specific reQTLs are common, both in *cis* and in *trans*, and that many *trans*-reQTLs likely function the same way in response to different stimuli. By comparing the responsiveness of genes and their underlying reQTLs across multiple stimuli, we refine known molecular circuits and infer the position of reQTLs within them.

## Results

### Step-by-step process to study reQTLs

We devised a general approach to study the genetic basis of variation in the transcriptional responsiveness to diverse stimuli (**Fig. 2a**). First, we isolate primary cells from genetically variable individuals (here, previously genotyped RI mouse strains), stimulate the cells *in vitro* with different stimuli (here, the TLR4 agonist lipopolysaccharide (LPS) which induces both inflammatory and anti-viral response genes; the TLR3 and MDA-5 agonist poly IC which induces anti-viral response genes; and the TLR2 agonist PAM which induces inflammatory response genes<sup>12, 13</sup>; **Fig. 2b**), and measure gene expression profiles in a relatively small number of individuals (here, 8 strains). Second, we use the gene expression profiles to define a 'variation signature' consisting of several hundred genes (here, 424 genes) whose responsiveness differs significantly between genetically distinct individuals. Third, we measure the variation signature (here, using the Nanostring nCounter) across samples from many individuals, under different stimuli and/or time points (here, 96 individuals, 3 stimuli and two time points). Fourth, we combine these signatures with genotypic information<sup>14</sup> to decipher the genetic architecture of variation in gene responsiveness. The initial genomic profiling allows us to determine a high-quality gene signature, while the higher throughput and low sample requirements of the subsequent signature assay allow scaling to a large number of stimuli, time points, and individuals, thus enhancing biological scope and statistical power.

### Variation in responsiveness of DCs from RI mice

We applied our step-wise approach to study the variation in transcriptional response of DCs from RI BXD mice<sup>11</sup>. We isolated DCs from individual age-synchronized female mice of six BXD strains and the two parental strains (B6, D2) and measured 30 global gene expression profiles in resting conditions and after stimulation with pathogen components (**Supplementary Table 1**). The profiles suggest significant genetic variation in the transcriptional response, with 40.6% of genes showing heritability that is higher than expected by chance (Wilcoxon test,  $P < 10^{-200}$ ; **Supplementary Note 1**).

We selected a 'variation signature' of 424 genes (Supplementary Note 2, Supplementary Tables 2 and 3), using a new method, InSignature, that selects genes based on genetic information so as to represent different underlying genetic factors regardless of their effect sizes. InSignature first partitions genes into distinct classes, each consisting of genes whose heritable variation in expression is likely associated with the same (shared) genetic factor (**Fig. 2c**), and then selects representatives from each class (**Fig. 2d**; **Supplementary Note 2**; **Methods**). InSignature performs well compared to other selection procedures by several measures (**Supplementary Fig. 1**).

Next, we measured the expression of the 424 signature genes in 384 samples of DCs from 96 individual mice: two individuals from each of 44 BXD strains and four individuals from each of the two parental strains (**Supplementary Table 4**). We measured mRNA abundance prior to and at 6 hours after LPS, polyI:C or PAM stimulation, using the nCounter (**Methods**). The nCounter measurements were highly consistent with those of microarrays (**Supplementary Fig. 2a-c**) and were highly reproducible (**Supplementary Fig. 2d**).

### Stimulus-specific *cis*-acting reQTLs

To identify reQTLs in different stimuli, we calculated — for each gene in each stimulus and strain — a responsiveness trait as the log ratio between that gene's expression level at 6 hours post stimulation and its baseline expression level (**Methods**). The heritability of responsiveness (percentage of responsiveness variance accounted for by strain) is much higher than expected by chance ( $P < 10^{-139}$ , permutation test). Thus, our data is amenable to genetic analysis of complex traits. We considered the responsiveness of each gene in each stimulus as an independent trait, and used a likelihood ratio (LR) score to determine the putative genetic variant(s) (reQTLs) that are significantly associated with each responsiveness trait per stimulus (**Methods**).

For 31 genes we detected a genetic association in *cis* in at least one stimulus (**Fig. 3a**; FDR-corrected association  $P$ -value  $< 0.01$ ; median heritability: 88%; **Methods**). Similar results are obtained from measuring the variation signature in LPS-stimulated DCs at an earlier (2h) time point, further supporting the identified *cis*-reQTLs (**Supplementary Note 3**). While each of the 31 genes is significantly *cis*-linked in one or two stimuli, those associations are insignificant in at least one other stimulus (**Fig. 3a**), even when lowering the significance threshold (**Methods**). Consistently, there is a significant gene-stimulus interaction in 28 of the 31 genes ( $P < 0.03$ ; **Supplementary Fig. 3a**). Thus, these *cis*-reQTLs are stimulus-dependent. This may parallel the common observation of different *cis*-eQTLs in different tissues<sup>15-18</sup>. The stimulus-dependency of *cis*-reQTLs is not merely a result of lack of responsiveness to some stimuli. While a high association (LR) score is related to higher (absolute) responsiveness in the gene (relative to that gene's responsiveness to other stimuli, **Supplementary Fig. 3b-e**), the reverse is not necessarily true (**Supplementary Fig. 3b,d**).

We partitioned the 31 *cis*-linked genes into eight groups (I-VIII, **Fig. 3a**) based on their level of responsiveness (**Fig. 3a**, grey scale) and *cis*-associations across stimuli (**Fig. 3a**, brown to yellow scale), respectively defining responsiveness and association profiles for each gene. The *cis*-reQTLs of the 19 genes in groups II, IV, VI, VII, VIII are stimulus-specific, since in at least one stimulus there is high absolute responsiveness but no genetic association. The remaining groups (I, III, and V, 12 genes) consist of stimulus non-specific *cis*-reQTLs, with matching *cis*-association and responsiveness profiles. On average, the stimulus non-specific *cis*-reQTLs have a larger effect size than stimulus-specific *cis*-reQTLs ( $P < 10^{-7}$ , Wilcoxon test; **Supplementary Fig. 3f**). This parallels previously observed differences between tissue non-specific and tissue-specific eQTLs<sup>19</sup>.

## InVamod identifies pleiotropic genetic variants

We next examined *trans*-acting reQTLs. Due to the lack of statistical power to detect *trans*-linkage using information on each trait independently<sup>20</sup> (**Supplementary Fig. 3g**), we focused on identifying pleiotropic variants that affect a set of co-varying genes. To this end, we devised InVamod, a method to identify pleiotropic genetic variants, their associated genes and the stimuli under which the association holds. InVamod takes as input all responsiveness traits and potential genotypes, partitions the traits into ‘co-variation modules’ and identifies for each module an associated genetic variant (**Fig. 4a, Supplementary Fig. 4, Methods**). InVamod first calculates the association scores for each trait at each genomic position, and uses this information to initialize a trivial seed solution of single-trait modules whose associated genetic variant is the best-scoring variant of the trait (**Fig. 4a, top**). It then agglomerates modules. In each agglomeration step, InVamod merges pairs of modules from the previous step into a new module, and updates the genetic variant of the new module such that it best reflects all traits in the newly generated module (**Fig. 4a, steps 1-3**). The updated variant might differ from the variants associated with the two agglomerated modules in the previous step (*e.g.*, variants *C* and *D* are replaced by variant *F*; **Fig. 4a, step 2**). Each merging step minimizes the loss (in association scores) due to updating the module's genetic variant. Merging (and updating) is only allowed if the loss is acceptable. The output is a set of modules, each consisting of traits and their (collectively) best associated genetic variant (**Fig. 4a, bottom**).

## Stimulus-specific pleiotropic *trans*-reQTLs

InVamod found six significantly large (possibly overlapping) modules in our data (**Fig. 4b,c, Supplementary Table 5**), each with at least 14 traits, the minimal size that would be expected by chance at a very low probability ( $P < 0.001$ , permutation test, **Supplementary Fig. 5a**). For example, the reQTL of module #2 (**Fig. 4b-d**) affects the responsiveness of the antiviral genes *Irf7*, *Stat2* and *Dhx58* under poly IC (a viral mimic), whereas the reQTL of module #5 (**Fig. 4b,c**) affects the responsiveness of the inflammatory genes *Cxcr4*, *NFkb2* and *Nlrp3* under PAM and LPS (bacterial components). Each module is linked to a distinct (*trans*) genomic position, and together spanning five different chromosomes with no particular location biases. Although the genomic intervals associated with modules #1, #2 and #4 include *cis*-associated genes (**Supplementary Table 6**), a potential source of propagated variation, these are not necessarily detected in the relevant conditions. The modules' *trans*-reQTLs explain on average 46% to 58% of the heritability of their genes (**Fig. 4b, Supplementary Fig. 5b,c, Methods**). Similarly to *cis*-reQTLs, a *trans*-reQTL implies high responsiveness of the cognate targets, but not vice versa (**Fig. 4d, Supplementary Fig. 5d**). The vast majority of *trans*-reQTLs are stimulus-specific (95% of modules' genes, **Supplementary Fig. 5d**), similar to the observed frequency of tissue specific *trans*-eQTLs (80-91% in Refs. <sup>16, 17, 21</sup>).

The traits in each module are significantly associated in coherent stimuli, forming a ‘reQTL activity profile’ (**Fig. 4b,d, Supplementary Fig. 5d; Methods**): mostly PAM in module #1, mostly poly IC in modules #2, #3 and #4, and PAM and LPS in modules #5 and #6 (**Fig. 4b,c**). In some cases, the same genes are members of more than one module, each associated

with different reQTLs in different stimuli (e.g., nine genes shared between module #1 and #2, **Fig. 4c**).

The transcriptional responsiveness of a module's genes is typically coordinated, likely reflecting their co-regulation. We define a gene's responsiveness profile as the set of stimuli to which the gene is highly responsive, and a module's 'leading responsiveness profile' based on significance of enrichment across the responsiveness profiles of its member genes (**Methods**). Five of the six modules have at least one leading (enriched) responsiveness profile (**Fig. 4b,d, Supplementary Fig. 5d**).

### Shared characteristics of association

Next, we aimed to decipher the mechanisms through which reQTLs affect target genes. In many cases, the same gene is associated with the same DNA region in response to more than one stimulus (*cis*: groups I-IV, **Fig. 3a**; *trans*: modules #5, #6; **Fig. 4c**). A parsimonious explanation for this observation is that the high responsiveness in these stimuli is mediated by the same mechanism, which is modulated by a single DNA polymorphism, for example by an effect on the binding of a regulatory protein that is active only during responses to those stimuli.

Two lines of evidence support this 'shared association, shared mechanism' hypothesis. First, the same reQTL is associated in combinations of stimuli that are consistent with the known architecture of the TLR circuit. *Trans*-reQTLs are associated in both PAM and LPS (modules #5 and #6, **Fig. 4b, Supplementary Fig. 5d**), suggesting interaction with inflammatory signaling pathways activated by both stimuli (**Fig. 2a**). *cis*-reQTLs are associated with both PAM and LPS (10 genes; groups I,II, **Fig. 3a**), which activate inflammatory transcription factors such as NF- $\kappa$ B (**Fig. 2a**), or with both poly IC and LPS (6 genes; groups III,IV, **Fig. 3a**), which activate antiviral factors such as members of the STAT and IRF families (**Fig. 2a**). Notably, no reQTLs were found associated with both PAM and poly IC; most known pathway components activated by these stimuli are distinct. Second, when responsiveness traits are classified into major classes based on their behavior across genotypes and time (**Fig. 3b**), a gene typically belongs to the same class in different stimuli (**Fig. 3a, Supplementary Fig. 5d**). Notably, we cannot completely rule out the non-parsimonious (and quite unlikely) possibility that due to the large linkage intervals, in some cases one reQTL in an interval may act in one stimulus and a neighboring reQTL may act in another, or that in the same stimulus, one reQTL in an interval may affect one gene and a neighboring reQTL may affect another gene.

### Positioning reQTLs in molecular circuits

Next, we exploited the shared characteristics of association and responsiveness profiles of a module's genes to devise InCircuit, a procedure to position a *trans*-reQTL in the context of a well-established molecular wiring map (a 'circuit'; **Fig. 5a**). InCircuit takes as input a (known) wiring diagram of a circuit associating stimuli to signaling pathways. First, InCircuit adds each of module gene at the bottom of the wiring map based on its responsiveness profile. Next, it positions the *trans*-reQTL associated with the module within the existing branches of the wiring diagram, if possible, such that (1) the stimuli upstream of

the reQTL match its predicted activity profile, and (2) the reQTL is upstream to a significant number of the target genes of the module. In particular, in this study, the reQTL is required to be upstream of all of its module's target genes, as long as they belong to a leading (enriched) responsiveness profile of the module. When more than one position is consistent with these two criteria, InCircuit selects the one that minimizes the number of 'irrelevant' (non-leading) downstream responsiveness profiles (**Supplementary Fig. 6, Methods**). *Trans*-acting reQTLs that can be positioned in this way are deemed consistent with the known circuit. Those that cannot may suggest new branches in the circuit (**Supplementary Fig. 6b**, 'model refinement'). This analysis can only be applied to modules with a least one enriched (leading) responsiveness profile (in our case, only five candidate reQTLs). In principle, the same rationale can be applied to *cis*-reQTLs (**Supplementary Fig. 7**). However, since *cis*-reQTL associations are determined from the expression of a single gene, they do not allow us to identify robustly enriched responsiveness profiles, and hence cannot be as reliably resolved in our data.

We applied InCircuit to a TLR wiring diagram that was previously constructed in DCs<sup>22</sup>. InCircuit first added genes at the bottom of the wiring map based on their responsiveness profiles (**Supplementary Fig. 5d, Methods**). The accuracy of InCircuit's differential positioning of genes downstream of branches 7, 8, 9 and 10 is strongly supported by their differential enrichment for binding by transcription factors from the RELA, STAT and IRF families, as measured by ChIP-Seq in DCs<sup>23</sup> (**Supplementary Fig. 8**). Next, InCircuit mapped each of four of the five candidate *trans*-reQTLs into a particular branch in the circuit diagram (**Fig. 5b**). For example, reQTL #5 is PAM- and LPS-specific and affects two enriched responsiveness profiles, PAM-LPS and PAM-LPS-poly IC. It was positioned in the branch propagating both PAM and LPS signaling, prior to the split into two distinct responsiveness profiles.

The fifth *trans*-reQTL, which is associated with module #1, could not be positioned in the known molecular circuit in a manner consistent with any existing branch. This module consists primarily of anti-viral genes (11 of 14,  $P < 10^{-7}$ , hypergeometric test, *e.g.*, Irf7, Stat2, Dhx58, Ifit2) with the LPS-poly IC (viral-like) responsiveness profile (**Fig. 4b**, 4<sup>th</sup> column) that are, however, *trans*-linked primarily under PAM stimulation (**Fig. 4b**, 3<sup>rd</sup> column; **Supplementary Fig. 5d**). To reconcile the reQTL with the circuit, we must add a potential cross-talk between bacterial and antiviral pathways (**Fig. 5b**, dashed line) and position the reQTL on this new branch. The existence of such a cross-talk has been previously postulated<sup>24, 25</sup>, and may be apparent only for some alleles.

Known interactions between transcription factors and the promoters of antiviral genes (**Fig. 5c**) provide support and insights into the potential bacterial-antiviral cross talk in module #1. In particular, promoter analysis<sup>26</sup> predicts that 9 of the 12 genes in module #1 are targets of the anti-viral IRF family of transcription factors ( $P < 10^{-7}$ , hypergeometric test; **Fig. 5c**, green rectangles in rightmost grey box), whereas only one gene is a predicted target of NF- $\kappa$ B (**Fig. 5c**, green rectangles in middle grey box). ChIP-Seq data indicates that the promoters of 8 of the 12 genes are bound by STAT1 in DCs<sup>23</sup> ( $P < 10^{-9}$ , hypergeometric test). This is consistent with their placement under branch 10. reQTL #1 may thus underlie genetic

variation in IRF or STAT activity, but only in a manner independent of NF- $\kappa$ B (**Fig. 5c**; branch 10).

### Identifying a gene underlying a reQTL

The *trans*-reQTL associated with module #2 is particularly intriguing because it affects the largest number of genes, its target genes are immune-related ( $P < 10^{-10}$ , **Fig. 4b**), and it is most significantly predicted to be active only during responses to poly IC ( $P < 10^{-16}$ ; **Fig. 4d**), but not LPS, PAM, or pre-stimulation (median LR score: 1.6 for eQTL).

To identify the gene likely underlying this *trans*-reQTL, we tested the effect of nine potential genes in the associated linkage interval at Chr1:129-169 Mb (**Fig. 6a**, **Supplementary Fig. 9a**) in stimulated DCs. These nine genes were selected from the 65 genes that reside within the linkage interval and are regulated in DCs during poly IC stimulation. Six genes (*Ikbke*, *Ptpcr*, *Ivns1abp*, *Fam129a*, *Rgs16*, *Rnasel*) have a non-synonymous coding polymorphism that alters the physicochemical group of the encoded amino acid, two genes have a *cis*-associated reQTL in either the array or the nCounter data (*Pla2g4a*, *Glul*; **Supplementary Table 6**), and *Rab2b* is a well-known direct effector of interferon production and hence a likely candidate.

We used shRNAs to knock down each candidate gene in primary bone marrow-derived DCs from B6 and D2 mice (one individual per strain), and measured the effect on our variation gene signature pre-infection and at 6 hours of infection with Sendai Virus (SeV). We used SeV because poly IC stimulation of shRNA-treated DCs lead to rapid cell death (data not shown), and SeV and poly IC affect the transcriptional response of key genes in similar ways<sup>27</sup>. Although poly IC and SeV are detected by distinct sensors (TLR3 and MDA-5 for poly IC, RIG-I and TLR7 for SeV<sup>28-30</sup>; **Fig. 5c**), both converge on the protein IPS-1 and its downstream signaling pathways, thus limiting any potential false positives to the short RIG-I-IPS branch.

*Rgs16* is the most likely candidate gene influencing module #2, since it was the only gene of all those tested with a robust and specific effect on the module ( $P < 10^{-24}$ ; **Fig. 6b**, **Methods**). This effect was observed with two different shRNAs targeting *Rgs16* and its magnitude was proportional to the degree of *Rgs16* knockdown (**Fig. 6b**). To determine if the *Rgs16* effect is specific to an anti-viral response, we compared the effect of shRNA specific for *Rgs16* on transcript levels 6 hours after stimulation with SeV infection, PAM, LPS or no stimulation (**Fig. 6c**, **Methods**). As predicted by our analysis, *Rgs16* shRNA significantly and specifically (t-test  $P$ -value  $< 10^{-3}$ ) reduced the expression of Module #2 genes after SeV infection, suggesting that *RGS16* acts as an activator of the module genes under viral stimulation. After PAM stimulation and in baseline (no stimulation) conditions, *Rgs16* shRNA had the opposite effect ( $P < 10^{-7}$ ), suggesting that *RGS16* acts as an inhibitor of the module in these conditions. After LPS stimulation, *Rgs16* shRNA had no effect on the module ( $P > 0.2$ ; **Fig. 6c**). The same results were obtained with two different shRNAs targeting *Rgs16* (**Fig. 6c** and **Supplementary Fig. 9b**), using both B6 and D2 mice (two biological replicates per strain; data not shown). One potential explanation for the effect of *Rgs16* activity is a single nucleotide polymorphism (SNP rs6362681) leading to a non-



synonymous amino acid change in residue 193 (glycine in B6, serine in D2)<sup>31</sup>, an unstructured serine-rich part of RGS16 that influences its signaling activity<sup>32</sup>.

## Discussion

Here, we present an experimental and computational framework to characterize the impact of genetic variation on transcriptional responsiveness to multiple stimuli, and we apply it to study the response of DCs to pathogen components. At the heart of our approach are two novel computational methods for multi-stimulus reQTL analysis: InVamod and InCircuit.

InVamod identifies *trans*-reQTLs with pleiotropic effects and characterizes their stimulus specificity. Previously published methods can jointly analyze multiple traits, but either do not scale well<sup>33,34, 35</sup> or generate results that are not directly biologically interpretable<sup>36</sup>. In contrast, InVamod analyzes multiple traits in an a-parametric, scalable manner, and produces coherent modules of significantly associated traits. Since InVamod takes as input the association scores rather than the trait values, it can in principle construct heterogeneous modules of distinct traits, allowing researchers to group traits regardless of their source and distribution.

InCircuit positions the discovered reQTLs within the branches of a known molecular circuit diagram. Circuit reconstruction methods position genetic variation in circuits based on gene expression data<sup>3, 37</sup>, but do not leverage the stimulus specificity of signaling pathways, reQTLs and target genes. In contrast, InCircuit directly uses stimulus-specificity of reQTLs (**Supplementary Fig. 10**), but only in the context of an established (prior) molecular network (**Fig. 5b**). If InCircuit cannot position a reQTL within the known circuitry, the reQTL can suggest new putative connections. InCircuit assumes that a single reQTL affects each responsiveness trait, and that the circuit diagram is deterministic and acyclic. Future developments that handle dynamics, feedback and uncertainty may increase the accuracy of reQTL positioning and improve our understanding of reQTL mechanism of action.

Our results lent strong support for RGS16 as the underlying cause of the module #2 reQTL. This regulator of G-protein coupled receptors is involved in inflammatory processes in B and T cells, and affects B cell function, leading to autoimmune disease in BXD2 mice<sup>38, 39</sup>. Rgs16 expression is induced in mouse and human DCs activated with LPS or poly IC<sup>40</sup>. Interestingly, RGS16 is contained within a copy number variation region that substantially differs within human Asian populations<sup>41</sup>.

To collect a large-scale, multi-stimulus dataset, we relied on a practical strategy that combines genome-wide profiling experiments, from which we define a ‘variation signature’, with a subsequent meso-scale assay to monitor the variation signature across many combinations of genotypes, stimuli and time points. Several different meso-scale technologies have cost, sensitivity and scale benefits (e.g., Luminex QuantiGene Plex, MRM<sup>42</sup>, Mass Cytometry<sup>43</sup>) that can be exploited as part of this design. While a meso-scale approach offers some benefits, genomic profiling may allow the identification of more variants (**Supplementary Fig. 1e**). As our overall framework is general, and InVamod and InCircuit are scalable and can be applied to multi-stimulus reQTL analysis with signature

data (as here) or genome-wide profiles (e.g., microarrays, RNA-seq), this approach may help decipher the stimulus specificity, molecular circuitry and underlying causes of a wide variety of reQTLs.

## Materials and Methods

### Preparation of dendritic cells

BMDCs were generated from 6- to 8-week-old female C57BL/6J, DBA/2J and BXD mice (Jackson Laboratories). All animal protocols were reviewed and approved by the MIT / Whitehead Institute / Broad Institute Committee on Animal Care (CAC protocol 0609-058-12). Bone marrow cells were collected from femora and tibiae and plated at  $10^6$  cells/ml on non-tissue culture treated Petri dishes in RPMI-1640 medium (GIBCO), supplemented with 10% FBS, L-glutamine, penicillin/streptomycin, MEM nonessential amino acids, HEPES, sodium pyruvate,  $\beta$ -mercaptoethanol, and murine GM-CSF (15 ng/ml; Peprotech) or human Flt3L (100 ng/ml; Peprotech). For microarray experiments, floating cells from GM-CSF cultures were sorted at day 5 by MACS using the CD11c (N418) MicroBeads kit (Miltenyi). Sorted CD11c<sup>+</sup> cells were used as GM-CSF-derived BMDCs, and plated at  $10^6$  cells/ml and stimulated at 16 hr post sorting. For all other experiments, GM-CSF-derived BMDCs were used directly and composed of ~90% CD11c<sup>+</sup> cells, which was comparable to sorted BMDCs and to our previous observations<sup>22</sup>. Notably, we excluded from this study those BXD strains that were recently found by genomic analysis to have unexpected heterozygosity contaminations (Fernando Pardo-Manuel de Villena, personal communication).

### Pathogen components and viruses

TLR ligands were from Invivogen (Pam3CSK4, LPS) and Enzo Life Sciences (poly IC), and were used at the following concentrations: Pam3CSK4 (250 ng/ml), poly IC (10  $\mu$ g/ml), LPS (100 ng/ml). Sendai virus was obtained from ATCC. DCs were infected with Sev in multiplicity of infection of 1.

### Microarray profiling and data preprocessing

All array hybridizations (**Supplementary Table 1**) and pre-processing were done using Affymetrix Mouse Genome 430A 2.0 Array as in Ref. <sup>22</sup> and normalized using RMA<sup>44</sup>. Probe sets that were absent in all samples according to Affymetrix flags were removed. Only 6,850 genes that have at least one measurement that is higher than a threshold value of 50 were used.

The standard log fold change is defined as log expression under stimulation, normalized by baseline log expression of the same strain. Inter-strain fold changes  $x/B6$  is the log expression in strain  $x$  normalized by the averaged log expression in strain B6 at the same stimulus and time point. We selected for further analysis 4,573 genes whose expression may be affected by either stimulus or strain, defined by  $\geq 1.7$  standard fold change or  $\geq 1.7$  inter-strain fold change.

## Selection of variation signature genes

The variation signature consists of 424 genes selected using four distinct criteria: **(I)** heritability in responsiveness, as scored by InSignature (below) (322 genes); **(II)** prior knowledge on a key role in antiviral and inflammatory responses or in related diseases, based on the annotation in the Ingenuity Knowledge Base (Ingenuity Systems, Mountain View, CA, USA) (61 genes); **(III)** positive responsiveness controls, based on DCs response (especially up regulation) to LPS, poly IC or PAM stimulation in previous report<sup>22</sup> (20 genes); and **(IV)** negative controls, based on low responsiveness variation (21 genes) (**Supplementary Note 2**).

Selection of genes based on heritable variation (criterion **(I)**) is challenging since (i) information should be combined across different stimuli, but only a few strains were globally profiled in each stimulus; and (ii) using top-scoring genes (*e.g.*, based on variance, heritability or association scores) might lead to an unbiased representation of the underlying genetic variants, over-representing target genes of variants with large pleiotropic effects. To address both challenges, we developed a new method, InSignature, as follows.

The InSignature method takes as input genome-wide profiles from different strains and stimuli and predicts, for each gene, an underlying ‘InSignature pattern’ comprising of a set of hypothetical optimal genetic variants that are likely to act in each of the stimuli. The InSignature patterns impose a partition of the genes into classes, each consisting of all genes with the same predicted InSignature pattern (**Supplementary Fig. 11**). InSignature then selects representative genes within each class. This reduces the dependence on effect size and multiplicity: genetic variants with strong influence on many genes will be represented similarly to variants with moderate or low effect on a few target genes only.

Formally, an InSignature pattern is a maximum likelihood fitting of expression traits into discrete values, as follows (**Supplementary Fig. 11**). For each gene, the method discretizes transcript abundances using a mixture of two or three Gaussians, so as to maximize a likelihood score. We further constrain the solution space so that individuals of the same strain and same condition must have the same discrete value. The InSignature score is the  $-\log_{10}$  P-value of a maximum likelihood ratio test, calculated as follows: The likelihood of the null hypothesis is calculated based on a single Gaussian, whereas the likelihood of the alternative hypothesis is the maximum likelihood fitting into two or three Gaussians. The test statistics is the logarithm of the likelihoods’ ratio, and its probability distribution is approximated using a chi-squared distribution (degrees of freedom are calculated based on Wilks’ theorem<sup>45</sup>) that allows the calculation of a P-value score. This approach allows **(1)** a principled filtering of genes; in our data, an InSignature score of 5.5 and higher corresponds to false discovery rate of 0.05 (**Supplementary Fig. 12**); **(2)** prediction of an underlying genotype in the case of multiple conditions and only a few strains, allowing a different genotype for each condition but requiring the same effect size; and **(3)** grouping together genes with the same InSignature pattern into ‘InSignature classes’ (**Supplementary Fig. 11**).

Next, to choose the final set of signature genes (criterion **(I)**), InSignature proceeds as follows. It first randomly selects a class and then chooses a gene within it. We use several alternative

criteria to select a representative gene within a chosen InSignature class: the top-scoring gene (based on the InSignature score); biologically relevant genes within the class (defined as criterion II above); responsiveness under LPS, poly IC and PAM (defined as criterion III above); and random selection of a gene within a class (full details are provided in **Supplementary Note 2**).

Here, of the 4,573 analyzed genes, 799 could be significantly fit with an InSignature score that is higher than 5.5 and were grouped into 446 InSignature classes. Of those, 294 are in singleton classes and the rest fall into 152 classes of size 2 to 13. Out of 322 genes selected by InSignature from these classes (typically one or two representatives per class), 232 were chosen as 'top scoring' in their class, 35 were randomly chosen from their class, 20 were chosen as class members that also have known biological relevance, and 35 were chosen as class members that are also highly responsive. To those we then added another 61 'biologically relevant' genes, 20 'responsive' genes and 20 negative controls, all of which were *not* within the classes (**Supplementary Table 3**).

### mRNA measurements on nCounter

The NanoString nCounter gene expression system captures and counts individual mRNA transcripts, using a multiplexed probe library (codeset) with sequence-specific probes for each gene of interest. Details on the nCounter system are presented in full in Ref. <sup>46</sup>. The nCounter codeset was constructed to detect each of 424 genes in the variation signature and used to measure mRNA expression on nCounter as previously described<sup>27</sup>.

### nCounter data preprocessing

We normalized the nCounter data (**Supplementary Table 7-10**) in three steps. First, we used positive spiked-in controls provided by the nCounter instrument to control for small variations in the efficiency of the automated process. Second, 21 control genes (**Supplementary Table 2**), which showed minimal variation based on the microarray data, were used to control for differences in sample values. Third, we removed covariates associated with epoch population substructure<sup>47</sup>. P-P plots demonstrating low observed inflation for representative trans-linked genes are presented in **Supplementary Fig. 13**.

### Association and variance components

Genotyping of 3,500 single nucleotide polymorphisms (SNPs) of BXDs were taken from WebQTL<sup>14</sup>. Responsiveness traits were calculated as the log ratio between the expression level under exposure to pathogen and the baseline expression level. (A responsiveness profile of a gene is the set of stimuli under which the gene is 'highly responsive', defined as exceeding 40% of the maximal measured responsiveness for that gene). Responsiveness-SNP associations were identified by assuming a two-level nested ANOVA model (higher level: genotypes; intermediate level: strains; lower level: individuals in each strain) and calculating its likelihood ratio (LR) statistic<sup>48</sup>. The 'LR score' is defined as  $2 \cdot \ln LR$ . The LR score indicates whether responsiveness differences could be related to the genetic marker rather than to differences between or within strains. To determine a significance threshold, we used a permutation test, reshuffling the BXD labels among strains, but performing no other reshuffling or randomization. In this way, we only disrupt the matching

between reQTLs and responsiveness but maintain within-strain variation. Using this permutation test, and requiring a false discovery rate of at most 0.001, we set the significance threshold for the LR score to be 29 for declaring a significant association of a single gene under a single stimulus. (For comparison, based on an approximation to an F statistic<sup>48</sup> and before multiple testing correction, an LR score of 29 corresponds to a type I error of  $3.5 \times 10^{-7}$ ). If the best scoring marker of a given responsiveness trait (or multiple traits) exceeds this threshold, it was termed a genetic variant (reQTL) associated with this trait.

Between-genotype and between-strain components of variance<sup>48</sup> were estimated based on the above two-level ANOVA model, allowing evaluation of (broad-sense) heritability ( $H^2$ ): the percentage of between-strain variance out of the overall variance. The percentage of heritability that is attributed to between-genotype effects is denoted % explained heritability<sup>49</sup>.

A responsiveness trait was classified based on comparing the degree of displacement (the difference between mean values of expression in strains carrying the two reQTL alleles) before and after stimulation (**Fig. 3b**). A trait is divergent if the (absolute) displacement is higher post-stimulation than at baseline. Divergence is early if the pre-stimulation displacement is already significant (here, 0.65 standard deviations difference), and late otherwise. Convergence indicates that the (absolute) displacement is lower post-stimulation than at baseline. This includes the special case of inverted traits (referred as ‘inversion’), when there is a significant displacement at both baseline and stimulated state, but at opposing directions.

To calculate interactions, we analyze the responsiveness data of the three stimulations by a full ANOVA model including the stimulation effect (S) and the reQTL×S interaction effect:

$$y_{ij} = \mu + S_j + Q_i + (Q \times S)_{ij} + e_{ij}$$

where  $y_{ij}$  is the gene's responsiveness at the  $i^{\text{th}}$  individual ( $i = 1, \dots, n$ ) and  $j^{\text{th}}$  stimulation;  $Q_i$  is the genotype effect;  $S_j$  is the  $j^{\text{th}}$  stimulation effect;  $e_{ij}$  is the residual error; and  $(Q \times S)_{ij}$  is the interaction effect between the  $i^{\text{th}}$  reQTL genotype and  $j^{\text{th}}$  stimulation.  $P$ -values for each of the effects are derived based on a corresponding F distribution as part of the full ANOVA test. For each gene, we first find the reQTL with the best LR score and then calculate the interaction effect for this reQTL.

### A scoring scheme for association of each single gene in multiple stimuli

The information from multiple traits of the same gene can provide an extra statistical power to detect significantly-linked genetic variants. We focused on multiple traits of the same gene that are linked to the same genetic variant. Such recurring event is highly unlikely to be generated at random, and therefore the LR score threshold can be reduced, while maintaining the same significance level. Because LR is an additive score, it is possible to calculate the association of multiple traits to a single genetic variant (marker) by summing over the LR scores of each of the traits alone. Based on empirical permutation tests, in our

dataset two traits must attain a total (summed) LR score of 46, that is, an average LR of only 23 for each trait, to provide an FDR-corrected P-value of 0.001 (compared to an LR of 29 for a single gene in a single stimulus, above). Altogether, we applied the following procedure. For each gene, we used a strict association score of 29 to detect genetic variants that act on single traits, or a permissive association score of 23 for cases of two or more traits that are best-linked to the same variant. We define an association profile of a gene and an reQTL as the set of stimuli under which the gene is significantly associated to that reQTL.

### Identification of co-variation modules (the InVamod procedure)

InVamod partitions the traits into co-variation modules in a step-wise agglomerative manner (**Fig. 4a, Supplementary Fig. 4**). Each module consists of a group of significant traits and the best scoring genetic variant for this group of traits. We define the association score between a module and a given variant as the sum of LR scores across all the module's traits to this variant. The best scoring variant of a given module is the variant that attains the highest association score for that module. This variant is typically different from the optimal variants that would be associated as the best-scoring marker of each of the traits independently (which may be different variants for different traits in the module). The association score of a module is therefore equal to or lower than the score of these optimal independent variants (denoted 'optimal score').

We initialize InVamod with the trivial seed solution of significant single-trait modules. At each step, InVamod merges traits or modules from the previous step, and then updates the best scoring genetic variant to the new module. The newly generated modules may attain a new genetic variant that is different from their variant at the previous step. A merge that results in a smaller association score loss (compared to the optimal score) is favored over those merges that cause a higher loss. The merging process is deterministic (optimizing the loss) and repeated until no two modules can be merged into a legal module. A module is legal if and only if **(1)** it is a multiplicative  $\epsilon$ -approximation of its optimal score; and **(2)** it does not contain traits that are insignificantly associated with its genetic variant. The cutoff for *trait* significance is predefined and depends on the number of traits within the module ('module size'): the larger the module, the lower the cutoff.

We evaluated module significance based on an empirical permutation test. Permutation tests were conducted by applying InVamod on 100 permuted datasets and identifying variation modules in each of these permuted datasets. InVamod was applied on the permuted datasets using the same parameters as applied on the original dataset (including  $\epsilon$  and LR cutoffs for traits in modules of sizes 1, 2 and 3; no other parameters are required). Each permuted dataset was generated by reshuffling each of the responsiveness traits independently. Reshuffling of a responsiveness trait was performed by changing responsiveness values among BXD strains while maintaining the original matching among responsiveness values that belong to two individuals of the same strain. This way, the original within-strain variation, between-strain variation and heritability is maintained, and only the association with a particular genetic variant is broken.

To detect significant modules (even when no single trait is highly significant), we constructed modules initially based on permissive cutoffs for trait significance and then applied stringent criteria to filter significantly large modules, based on permutation tests. In our analysis, we used  $\epsilon=0.7$ , and permissive cutoffs of LR = 16, 10, 9 for modules of size 1, 2, 3 and larger, respectively (for comparison, LR of 29 and 23 provide stringent cutoffs for one or two traits as detailed above). Based on empirical permutation tests, with these parameters, module size of 14 and higher is unlikely to be generated by chance (P-value 0.01; **Supplementary Fig. 5a**).

### Positioning *trans*-reQTLs on a regulatory circuit (the InCircuit procedure)

InCircuit takes as input a co-variation module and a wiring diagram (denoted ‘regulatory circuit’) depicting all relevant stimuli, signaling pathways, and responsiveness profiles (RPs) under study. The reQTL activity profile of a module is the collection of stimuli under which the reQTL is predicted to be active (predicted activity under a given combination of stimuli is based on over-representation of module traits under these particular stimuli; FDR-corrected hyper-geometric P-value  $<10^{-3}$ ). Over-represented RPs of a module are its ‘leading RPs’. A *trans*-acting reQTL is added to a branch on the wiring diagram based on two criteria: the stimuli upstream the reQTL exactly match its activity profile, while all the module's leading RPs are downstream to it (**Supplementary Fig. 6a**). When more than one possible branch satisfies these conditions, a model selection criterion is applied: the reQTL position is selected such that it minimizes the total number of downstream (leading or non-leading) RPs. When no such branch exists, a new branch is added to the wiring diagram consistently with these requirements (**Supplementary Fig. 6b**, ‘model refinement’).

InCircuit's input regulatory circuit (**Fig. 5b**) was not constructed as part of this study but was taken as prior knowledge from a comprehensive previous study of the transcriptional response to TLR stimulation in DCs<sup>22</sup>, and is consistent with extensive literature on the DC TLR circuit<sup>13</sup>. The target genes were added to the diagram based on their measured responsiveness profiles (**Supplementary Fig. 5d**). The wiring diagram in **Fig. 5c** represents protein-protein interactions map of TLR signaling from the Ingenuity Knowledge Base (Ingenuity Systems, Mountain View, CA, USA), where the target genes were added based on a computational promoter analysis<sup>26</sup>.

### shRNA knockdowns

High-titer lentiviruses encoding shRNAs targeting genes of interest were obtained from The RNAi Consortium (TRC; Broad Institute, Cambridge, MA, USA)<sup>50</sup>. Bone marrow cells were infected with lentiviruses as previously described<sup>22</sup>. For each gene of interest, we verified knock down efficiency in our cells using qPCR of the target gene. To this end, we harvested PolyA+ RNA in 96-well plates with the Turbocapture mRNA kit (Qiagen) and reverse transcribed with the Sensiscript RT kit (Qiagen). Real time quantitative PCR reactions were performed on the LightCycler 480 system (Roche) with FastStart Universal SYBR Green Master Mix (Roche). Every reaction was run in triplicate and GAPDH levels were used as an endogenous control for normalization. All shRNA used in the study had >50% knockdown efficiency. The two shRNAs targeting Rgs16 (shRGS16(1) and shRGS16(2), **Fig. 6**) have distinct seed region sequences (1:

GCCAGAACTTATACTTTCTA; 2: CCGAGAACTGACCAAGACAAA). Lentivirus-infected cells were composed of ~90% CD11c<sup>+</sup> cells, which was comparable to sorted BMDCs and to our previous observations<sup>22</sup>. The effect of shRNA knockdown on gene expression was quantified using the nCounter assay on shRNA-treated DCs before and 6 hours after stimulation with PAM, LPS or infection with SeV at MOI of 1.

The shRNA effect on the responsiveness each gene is defined as the difference between responsiveness of the shRNA-treated cells compared to responsiveness of LacZ shRNA and Luciferase shRNA control-treated cells (responsiveness is averaged across B6 and D2 in both cases). Similarly, the shRNA effect on transcript level of each gene is defined as log<sub>2</sub> (expression level in shRNA-treated cells / expression level of LacZ shRNA and Luciferase shRNA control-treated cells). The preferential effect of a reQTL on the responsiveness of the module genes is the -log P-value of a *t*-test for the difference between the distribution of shRNA effects on responsiveness of module genes compared to the distribution of effects on the remaining genes in the assay.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We thank Chun (Jimmie) Ye, Joseph Pickrel, Mark Daly and Eric Lander for comments and discussions. IGV and IA were supported by Human Frontiers Science Program postdoctoral fellowships. Work was supported by HHMI, an NIH PIONEER award, a Burroughs-Wellcome Fund Career Award at the Scientific Interface (AR), a Center for Excellence in Genome Science grant 5P50HG006193-02 from the NHGRI (NH and AR), the Klarman Cell Observatory at the Broad Institute (AR), the New England Regional Center for Excellence/Biodefense and Emerging Infectious Disease U54 AI057159 (NH), the Israeli Centers of Research Excellence (I-CORE) Gene Regulation in Complex Human Disease, Center No 41/11 (IGV), the Human Frontiers Science Program Career Development Award and an ISF Bikura Institutional Research Grant Program (IA) and the Edmond J. Safra Center for Bioinformatics at Tel-Aviv university (RW and YS). AR is a fellow of the Merkin Foundation for Stem Cell Research at the Broad Institute. IGV is a Faculty Fellow of the Edmond J. Safra Center for Bioinformatics at Tel Aviv University and an Alon Fellow.

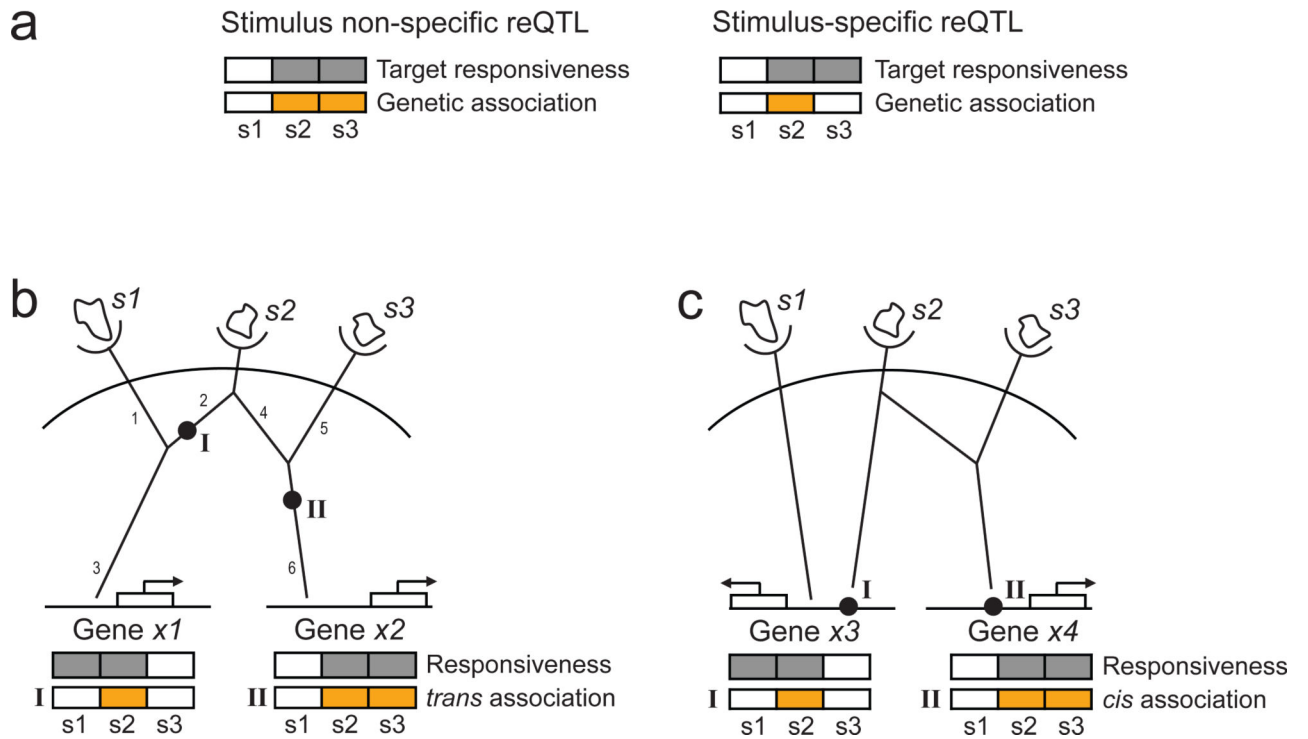
## References

1. Smith EN, Kruglyak L. Gene-environment interaction in yeast gene expression. *PLoS Biol.* 2008; 6:e83. [PubMed: 18416601]
2. Heinig M, et al. A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. *Nature.* 2010; 467:460–464. [PubMed: 20827270]
3. Zhong H, et al. Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. *PLoS Genet.* 2010; 6:e1000932. [PubMed: 20463879]
4. Barreiro LB, et al. Deciphering the genetic architecture of variation in the immune response to *Mycobacterium tuberculosis* infection. *Proc Natl Acad Sci U S A.* 2012; 109:1204–1209. [PubMed: 22233810]
5. Gargalovic PS, et al. Identification of inflammatory gene modules based on variations of human endothelial cell responses to oxidized lipids. *Proc Natl Acad Sci U S A.* 2006; 103:12741–12746. [PubMed: 16912112]
6. Smirnov DA, Morley M, Shin E, Spielman RS, Cheung VG. Genetic analysis of radiation-induced changes in human gene expression. *Nature.* 2009; 459:587–591. [PubMed: 19349959]
7. Yang IV, et al. Identification of novel genes that mediate innate immunity using inbred mice. *Genetics.* 2009; 183:1535–1544. [PubMed: 19805818]



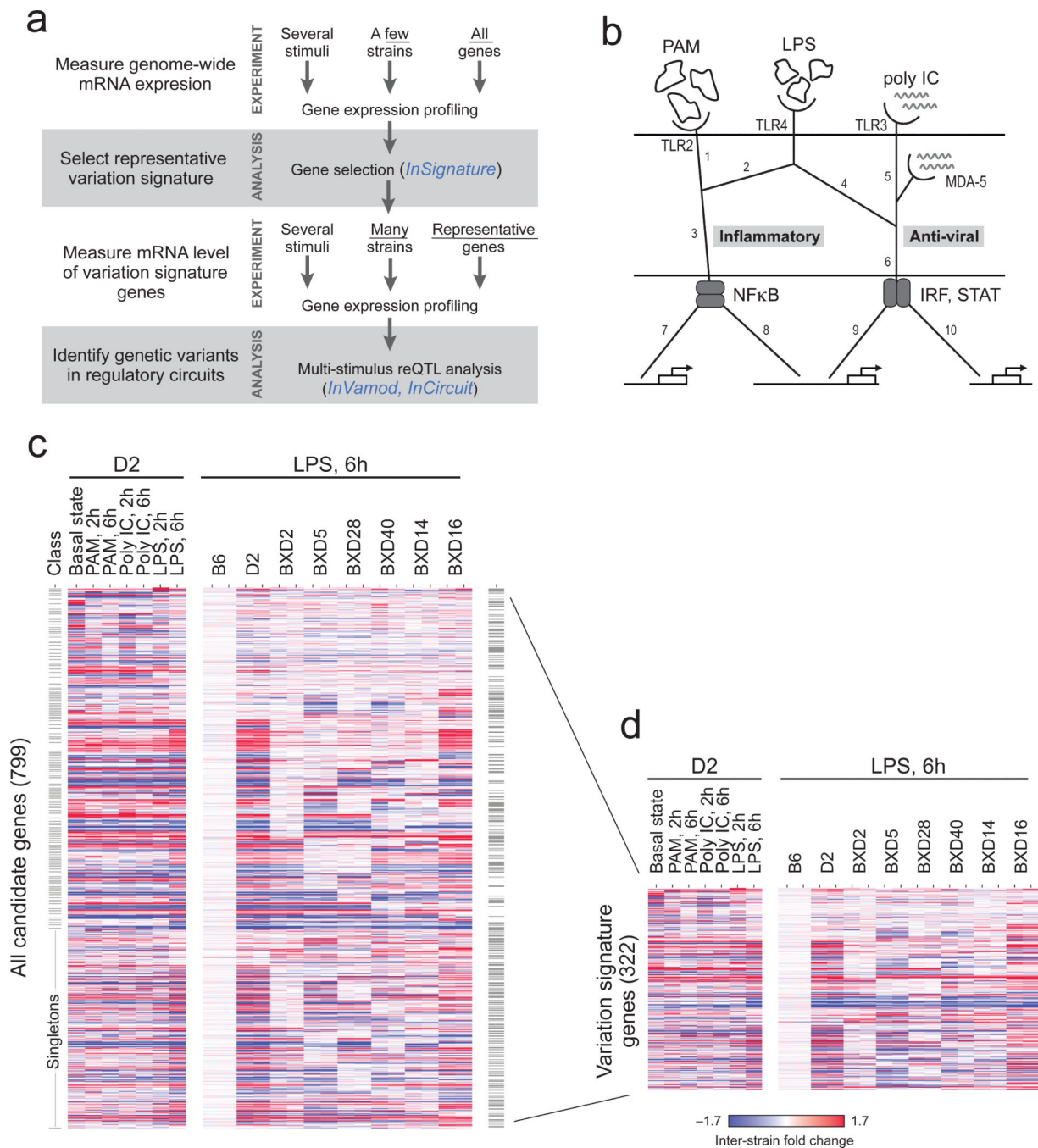
8. Dombroski BA, et al. Gene expression and genetic variation in response to endoplasmic reticulum stress in human cells. *Am J Hum Genet.* 2010; 86:719–729. [PubMed: 20398888]
9. Romanoski CE, et al. Systems genetics analysis of gene-by-environment interactions in human cells. *Am J Hum Genet.* 2010; 86:399–410. [PubMed: 20170901]
10. Maranville JC, et al. Interactions between glucocorticoid treatment and cis-regulatory polymorphisms contribute to cellular response phenotypes. *PLoS Genet.* 2011; 7:e1002162. [PubMed: 21750684]
11. Peirce JL, Lu L, Gu J, Silver LM, Williams RW. A new set of BXD recombinant inbred lines from advanced intercross populations in mice. *BMC genetics.* 2004; 5:7. [PubMed: 15117419]
12. Takeda K, Akira S. Toll-like receptors in innate immunity. *Int Immunol.* 2005; 17:1–14. [PubMed: 15585605]
13. Kawai T, Akira S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nature immunology.* 2010; 11:373–384. [PubMed: 20404851]
14. Wang J, Williams RW, Manly KF. WebQTL: web-based complex trait analysis. *Neuroinformatics.* 2003; 1:299–308. [PubMed: 15043217]
15. Chesler EJ, et al. Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nat Genet.* 2005; 37:233–242. [PubMed: 15711545]
16. Bystrykh L, et al. Uncovering regulatory pathways that affect hematopoietic stem cell function using ‘genetical genomics’. *Nat Genet.* 2005; 37:225–232. [PubMed: 15711547]
17. Fairfax BP, et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat Genet.* 2012; 44:502–510. [PubMed: 22446964]
18. Gerrits A, et al. Expression quantitative trait loci are highly sensitive to cellular differentiation state. *PLoS Genet.* 2009; 5:e1000692. [PubMed: 19834560]
19. Dimas AS, et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science.* 2009; 325:1246–1250. [PubMed: 19644074]
20. Mackay TF, Stone EA, Ayroles JF. The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet.* 2009; 10:565–577. [PubMed: 19584810]
21. Hubner N, et al. Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat Genet.* 2005; 37:243–253. [PubMed: 15711544]
22. Amit I, et al. Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science.* 2009; 326:257–263. [PubMed: 19729616]
23. Garber M, et al. A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Molecular cell.* 2012; 47:810–822. [PubMed: 22940246]
24. O’Neill LA. When signaling pathways collide: positive and negative regulation of toll-like receptor signal transduction. *Immunity.* 2008; 29:12–20. [PubMed: 18631453]
25. Barbalat R, Lau L, Locksley RM, Barton GM. Toll-like receptor 2 on inflammatory monocytes induces type I interferon in response to viral but not bacterial ligands. *Nature immunology.* 2009; 10:1200–1207. [PubMed: 19801985]
26. Ulitsky I, et al. Expander: from expression microarrays to networks and functions. *Nature protocols.* 2010; 5:303–322. [PubMed: 20134430]
27. Chevrier N, et al. Systematic discovery of TLR signaling components delineates viral-sensing circuits. *Cell.* 2011; 147:853–867. [PubMed: 22078882]
28. Baum A, Garcia-Sastre A. Differential recognition of viral RNA by RIG-I. *Virulence.* 2011; 2:166–169. [PubMed: 21422808]
29. Diebold SS, Kaisho T, Hemmi H, Akira S, Reis e Sousa C. Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. *Science.* 2004; 303:1529–1531. [PubMed: 14976261]
30. Kato H, et al. Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature.* 2006; 441:101–105. [PubMed: 16625202]
31. Frazer KA, et al. Segmental phylogenetic relationships of inbred mouse strains revealed by fine-scale analysis of sequence variation across 4.6 mb of mouse genome. *Genome Res.* 2004; 14:1493–1500. [PubMed: 15289472]

32. Chen C, Wang H, Fong CW, Lin SC. Multiple phosphorylation sites in RGS16 differentially modulate its GAP activity. *FEBS letters*. 2001; 504:16–22. [PubMed: 11522288]
33. Hackett CA, Meyer RC, Thomas WT. Multi-trait QTL mapping in barley using multivariate regression. *Genetical research*. 2001; 77:95–106. [PubMed: 11279835]
34. Xu C, Li Z, Xu S. Joint mapping of quantitative trait Loci for multiple binary characters. *Genetics*. 2005; 169:1045–1059. [PubMed: 15489509]
35. Banerjee S, Yandell BS, Yi N. Bayesian quantitative trait loci mapping for multiple traits. *Genetics*. 2008; 179:2275–2289. [PubMed: 18689903]
36. Gilbert H, Le Roy P. Methods for the detection of multiple linked QTL applied to a mixture of full and half sib families. *Genetics, selection, evolution : GSE*. 2007; 39:139–158.
37. Lee SI, Pe'er D, Dudley AM, Church GM, Koller D. Identifying regulatory mechanisms using individual variation reveals key role for chromatin modification. *Proc Natl Acad Sci U S A*. 2006; 103:14062–14067. [PubMed: 16968785]
38. Xie S, et al. IL-17 activates the canonical NF-kappaB signaling pathway in autoimmune B cells of BXD2 mice to upregulate the expression of regulators of G-protein signaling 16. *J Immunol*. 2010; 184:2289–2296. [PubMed: 20139273]
39. Shankar SP, et al. RGS16 Attenuates Pulmonary Th2/Th17 Inflammatory Responses. *J Immunol*. 2012; 188:6347–6356. [PubMed: 22593615]
40. Shi GX, Harrison K, Han SB, Moratz C, Kehrl JH. Toll-like receptor signaling alters the expression of regulator of G protein signaling proteins in dendritic cells: implications for G protein-coupled receptor signaling. *J Immunol*. 2004; 172:5175–5184. [PubMed: 15100254]
41. Park H, et al. Discovery of common Asian copy number variants using integrated high-resolution array CGH and massively parallel DNA sequencing. *Nat Genet*. 2010; 42:400–405. [PubMed: 20364138]
42. Addona TA, et al. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat Biotechnol*. 2009; 27:633–641. [PubMed: 19561596]
43. Bendall SC, et al. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science*. 2011; 332:687–696. [PubMed: 21551058]
44. Irizarry RA, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003; 4:249–264. [PubMed: 12925520]
45. Wilks SS. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics*. 1938; 9:60–62.
46. Geiss GK, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol*. 2008; 26:317–325. [PubMed: 18278033]
47. Gaglani SM, Lu L, Williams RW, Rosen GD. The genetic control of neocortex volume and covariation with neocortical gene expression in mice. *BMC neuroscience*. 2009; 10:44. [PubMed: 19426526]
48. scheffe, H. *The analysis of variance*. John Wiley and Sons, Inc.; 1959.
49. Falconer, D.; Mackay, TFC. *Introduction to Quantitative Genetics*. Pearson educational limited; 1996.
50. Root DE, Hacohen N, Hahn WC, Lander ES, Sabatini DM. Genome-scale loss-of-function screening with a lentiviral RNAi library. *Nature methods*. 2006; 3:715–719. [PubMed: 16929317]



**Figure 1. Multi-stimulus reQTL analysis**

(a) Shown are responsiveness traits (grey) and genetic associations (orange) for stimulus non-specific (left) and stimulus-specific (right) reQTLs in each of three stimuli ( $s_1$ ,  $s_2$ ,  $s_3$ ). Stimulus-specificity is determined by comparing the stimuli in which the gene is responsive (grey) to the stimuli in which this responsiveness is associated with the genetic variants (orange). (b-c) Shown is a conceptual scheme for relating genetic variants (black filled circles) to branches in molecular circuits (black lines) in the context of multiple stimuli ( $s_1$ ,  $s_2$ ,  $s_3$ ), based on considering the stimuli in which we observe responsiveness traits (grey) in genes ( $x_1$ ,  $x_2$ ,  $x_3$ ,  $x_4$ ) and the stimuli in which those traits are associated with the genetic variants (orange). (b) Positioning *trans*-reQTLs. The position of reQTL **I** in branch 2 is consistent with its activity only in stimulus  $s_2$  and with its effect on gene  $x_1$ , which responds to both  $s_1$  and  $s_2$ . The position of reQTL **II** in branch 6 is consistent with its activity in both stimuli  $s_2$  and  $s_3$  and with its effect on gene  $x_2$ , which responds to both stimuli. (c) Positioning *cis*-reQTLs. The *cis*-reQTL **I** interacts with factors from a branch stimulated by  $s_2$  only, whereas *cis*-reQTL **II** responds to factors affected by both  $s_2$  and  $s_3$ .



**Figure 2. Experimental design of the study in DCs**

(a) Overview of step-wise study design applied here to the TLR circuit in DCs. Blue text indicates new computational methods. (b) A schematic map of the TLR circuit under study, consisting of three agonists (PAM, poly IC and LPS), their cognate receptors (TLR2, TLR3 and MDA5, and TLR4), branches propagating the signal through transcription factors such as NF-κB and Irf3 or STAT (rounded gray rectangles) and downstream target genes (white rectangles, bottom). (c) Shown is the expression of 799 significantly scoring genes (rows) in DCs from the D2 parental strain (left) and in each parental and 6 inbred BXD strains (right)

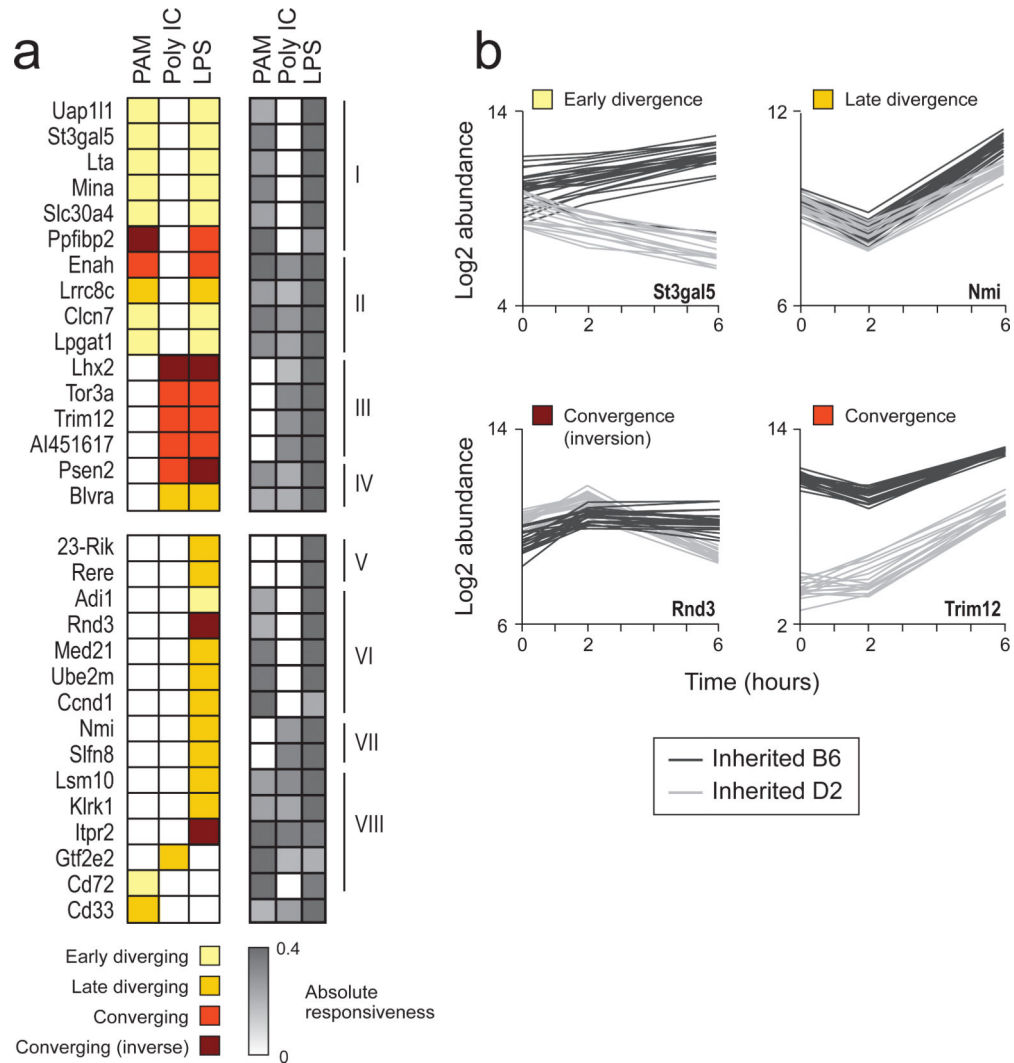
after stimulation for the indicated time periods with pathogen components; 2 individuals of each strain are shown consecutively. Expression values are log fold changes relative to the B6 strain at the same condition (red/blue color scale); a total of 30 arrays were used (**Supplementary Table 1**). Genes were clustered into distinct classes based on the InSignature algorithm; class separation is indicated by horizontal lines (left column); singleton classes are labeled. **(d)** Same as (c) but showing only the 322 representative genes chosen by InSignature (out of the 799 genes in (c)) to be included in the variation signature.

Author Manuscript

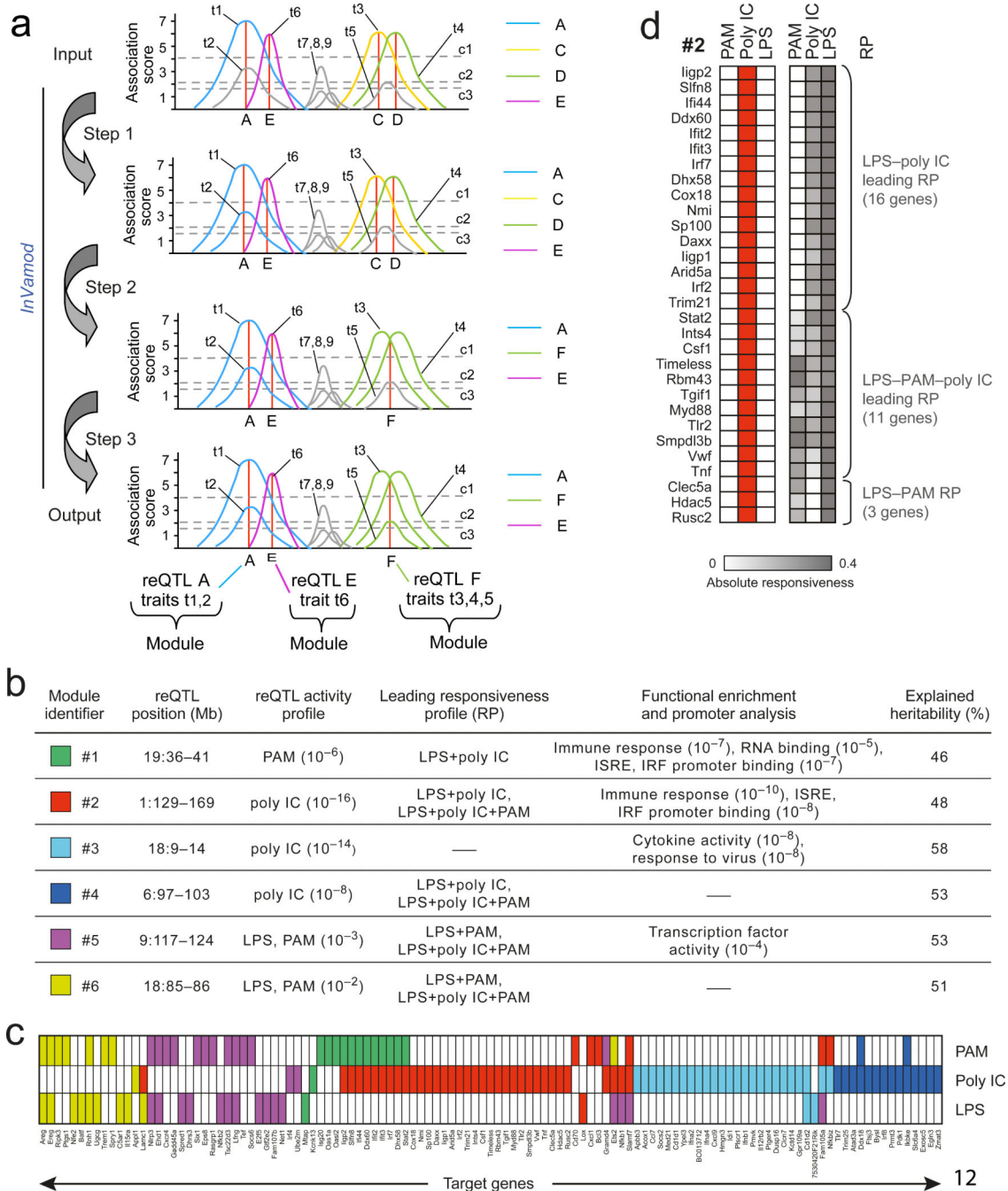
Author Manuscript

Author Manuscript

Author Manuscript



**Figure 3. cis-reQTLs in the response of DCs to three pathogenic components**  
**(a)** Shown are heatmaps of *cis*-association profiles (left; high likelihood ratio (LR) scores - colored; low LR scores – white) and of responsiveness profiles (right; dark grey: high responsiveness, white: low responsiveness) for all genes (rows) that are *cis*-associated in at least one stimulus (columns). The genes are split into eight groups (I-VIII) based on the combination of responsiveness and association profiles. **(b)** Shown are expression profiles (y axis) at three time points following LPS stimulation (x axis) of genes representing different association classes: early divergence (St3gal5), late divergence (Nmi), convergence (Trim12) and inversion (Rnd3). Depicted are profiles of B6, D2, and 44 BXD strains; Black and grey denote whether the strain carries B6 or D2 alleles in the proximity of the gene of interest.

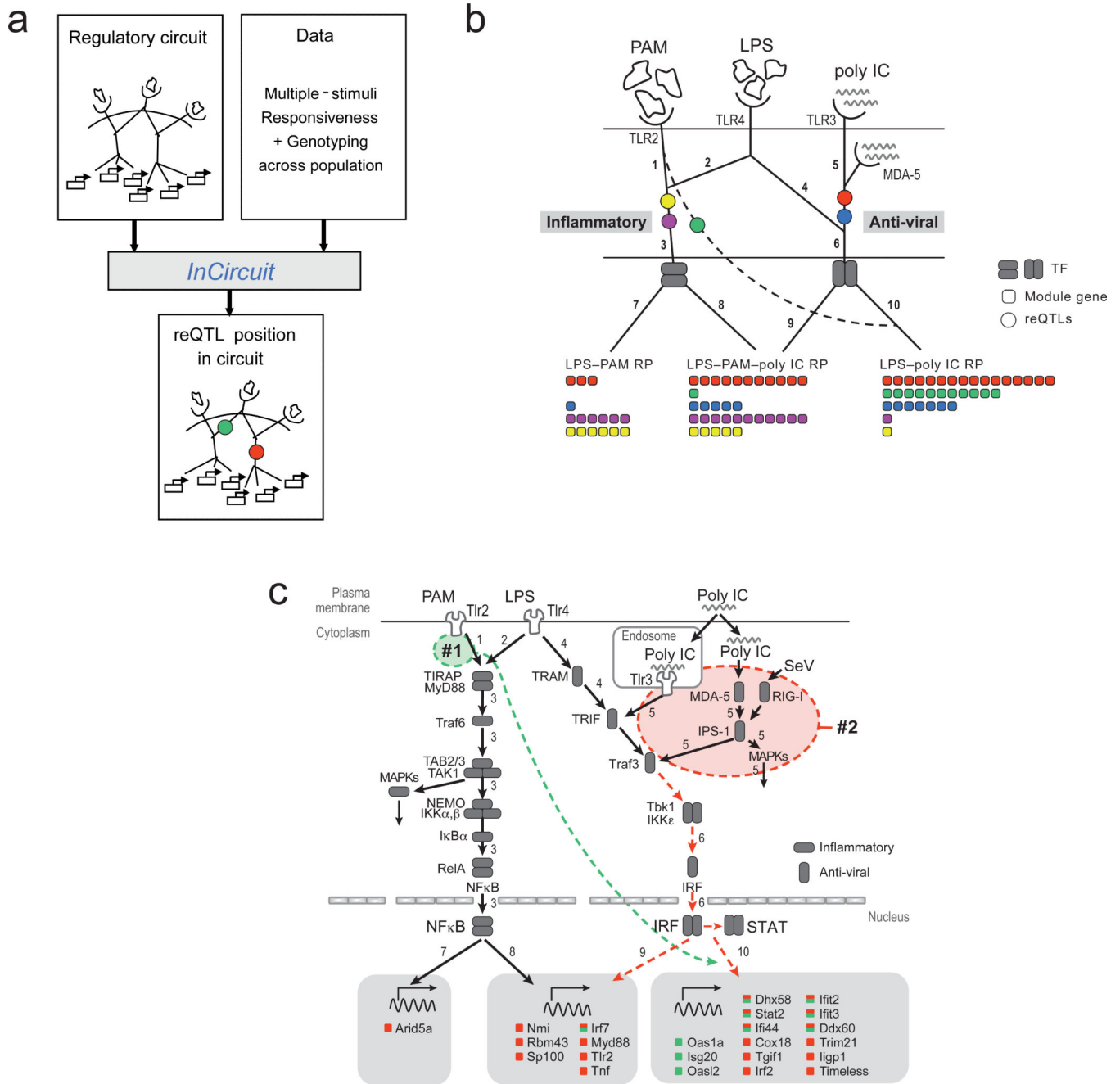


**Figure 4. Stimulus-specific pleiotropic trans-acting reQTLs**

(a) Illustration of the InVamod hierarchical reconstruction procedure for *trans*-reQTL analysis. InVamod starts (top) with association scores (y-axis) of traits (t1–t9, marked) at each genomic position (x-axis). Dashed lines are significance cutoffs for single, double and triple trait modules (c1=4, c2=2, c3=1.5, respectively). First, InVamod initializes modules (top): t1, t3, t4, and t6 form four seed modules associated at positions A, C, D and E, respectively. Other traits do not exceed c1 at any position. Each module is denoted by its associated genetic variant (vertical red lines) and is color-coded (A-blue, C-orange, D-green,

E-purple). Next, InVamod constructs modules hierarchically by merging traits and updating the genetic variants to the new module (**Methods**). In this illustrative example in step 1: Module A and t2 merged and the new module remains associated at A (as per **Supplementary Fig. 4a**); in step 2: Modules C and D merged (as per **Supplementary Fig. 4b**) and associated with a new genetic variant F; in step 3: Module F and t5 merged, and the merged module remains associated at F. Other traits, such as t7, t8, t9 and Module E cannot be legally merged (as per **Supplementary Fig. 4c**). InVamod outputs (bottom) three modules consisting of 2, 1 and 3 traits that are associated with genetic variants A, E, and F, respectively. **(b)** Co-variation modules. Shown are key characteristics for the six significant modules identified by InVamod. **(c)** Association profiles of genes (columns) in each variation module. Module association is indicated by the color code in column 1 in **(b)**. *cis*-linked traits are omitted. **(d)** Shown are the association profiles (left, red denotes associated) of the genes (rows) in module #2 and their responsiveness profiles (right, grey denotes responsiveness). Distinct responsiveness profiles (RPs) are marked on the right. Genes that are not associated under poly IC or are associated with additional modules are excluded for simplicity. Similar presentation of the other five modules is in **Supplementary Fig. 5d**.





**Figure 5. Positioning reQTLs in regulatory circuits**

(a) Illustration of the InCircuit algorithm. For details see **Supplementary Fig. 6**. (b) InCircuit's positioning five reQTLs (colored circles) in the TLR molecular circuit. Solid edges: branches of the input circuit, adapted from previous reports<sup>22</sup>. Target genes of each module (filled colored squares at bottom, color coded as in **Fig. 4b**) are arranged by their responsiveness profiles (as detailed in **Supplementary Fig. 5d**) downstream of transcription factors (TF, rounded gray rectangles). Dashed edge: new branch required to reconcile our data with the existing molecular circuit. (c) Shown is a signaling network of a TLR circuit (adapted from previous reports<sup>12, 13</sup>) with the predicted positions of the *trans*-reQTLs

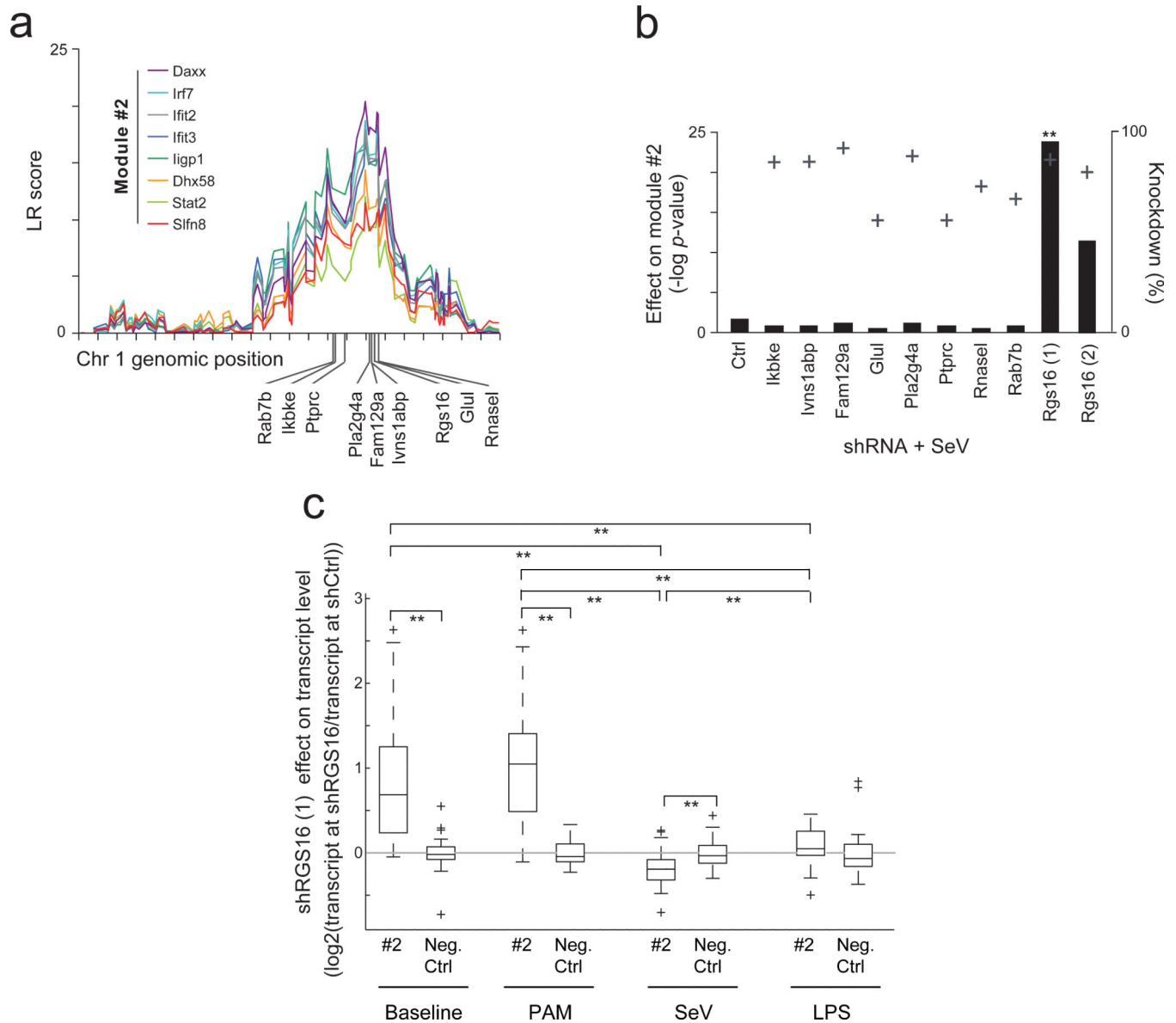
for module #1 (green oval) and #2 (red oval) based on InCircuit's analysis of (b). Module #1 and #2 genes are positioned under either IRF/STAT or NF- $\kappa$ B based on promoter analysis<sup>26</sup>, and color-coded as in **Fig. 4b**. If a gene is a member of both module #1 and #2 it is indicated with both colors. Dashed colored arrows connect each reQTL to its target genes.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 6. Rgs16 may be the causal variant in the module #2 reQTL**  
**(a)** Module #2 reQTL in Chr1: 129-165Mb. Shown is the LR score (Y axis) across the chromosome 1 genomic position (x-axis) for LPS responsiveness traits in eight module #2 genes. The positions of the nine candidate causal variants in this interval are marked below the x-axis. **(b)** Effect of shRNAs specific for each of the nine candidate causal variants (2 shRNAs for Rgs16 were used) or control shRNA on responsiveness of module #2 genes in SeV-infected DCs. Y-axis indicates the -log P-value of t-test (**Methods**) (black bars, left y-axis), as well as the extent of knockdown (+ marks, right y-axis) caused by each shRNA. **(c)** Effect of Rgs16 shRNA #1 on abundance of transcripts encoding module #2 genes or negative control genes (log<sub>2</sub>(transcript level in Rgs16 shRNA vs. its level in Luciferase and LacZ shRNA), averaged across a B6 and D2 individuals; **Methods**) in different stimulation conditions (6h after SeV infection, PAM, or LPS and no stimulation (baseline)). Positive/negative values indicate increased/decreased transcript levels following Rgs16 shRNA

treatment. In each box, the central mark is the median; edges are the 25th and 75th percentiles; whiskers extend to the most extreme data points not considered outliers; and outliers are plotted individually. \*\* $P < 10^{-3}$ .

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript